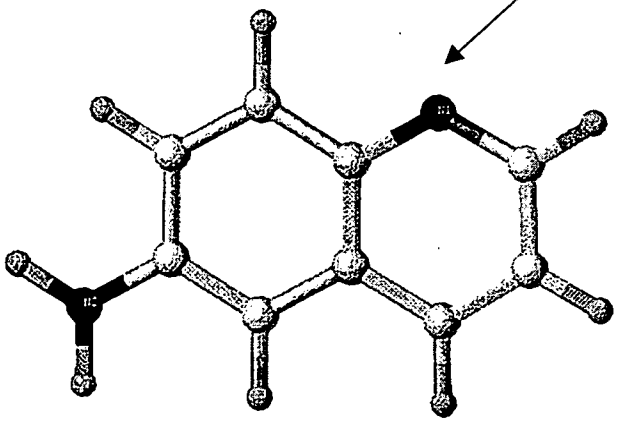
FIGURE 1

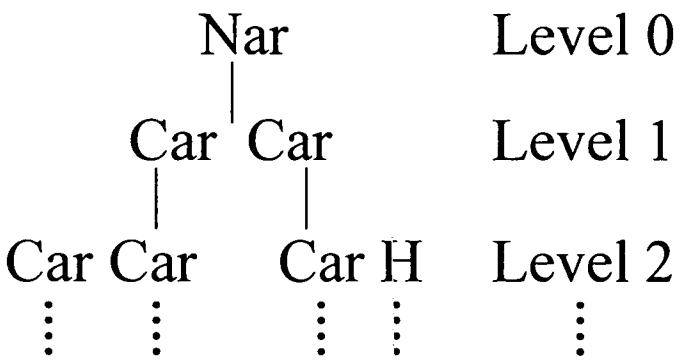
Environment Around an Ionizable Center

- **E.g. 6-aminoquinoline**

Measured 5.6
predicted 5.0



Start with interesting atom
find connections
find connections to connections
create a tree down to 5 levels
'bin' the atom types for each level
create a 'fingerprint' for this atom



String contains a bin for each required atom type at each level, the number of atom types is accumulated to form the string

FIGURE 3

Environment Around an Ionizable Center

- In addition to atom types, certain group types are also used and improve the performance of the algorithm

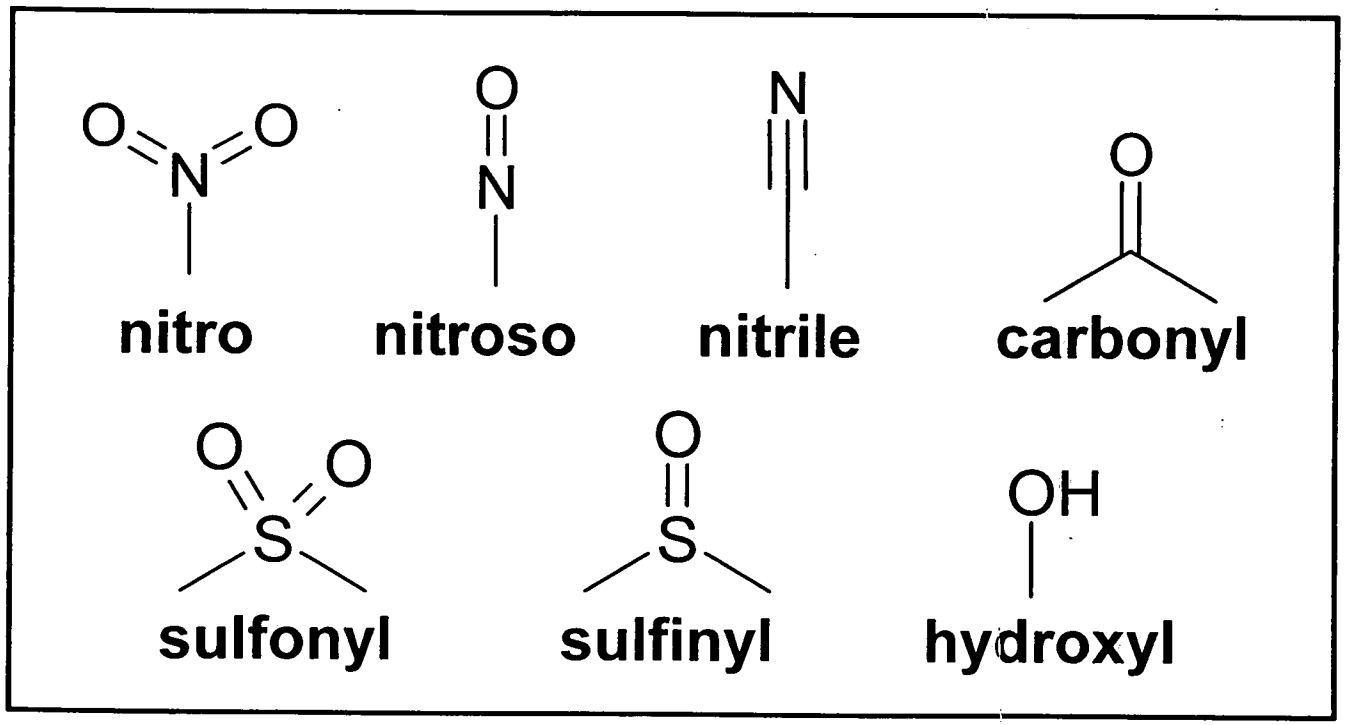
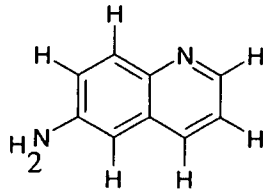


FIGURE 4



Atom Type:	C.3	C.2	C.1	C.ar	C.cat	N.3	N.2	N. 1	N.ar	N.am	N.pl3	N.4	O.3	0.2	0.CO2	S.3	S.2	S.O	S.O2	P.3	H	F	CL	BR	I	NO2	NO	CN	CO	SO2	SO	OH
Level 0									1																							
Level 1				2																												
Level 2				3																	1											
Level 3				3																	2											
Level 4				1																	3											
Level 5											1																					

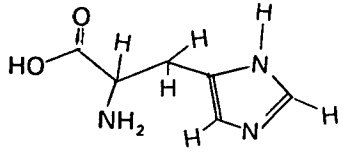
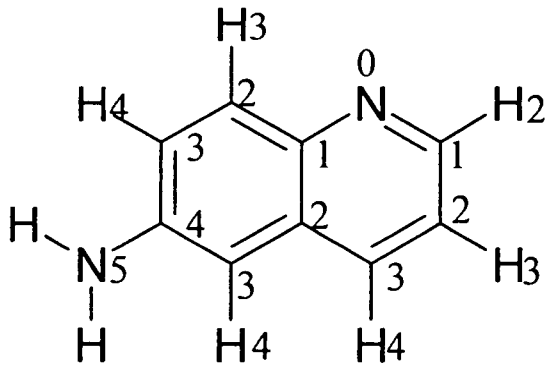
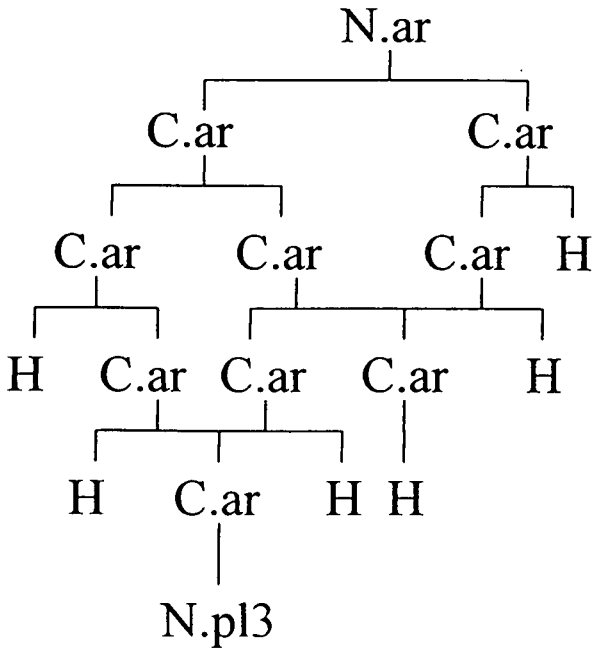


FIGURE 5

Atom Type: .	C.3	C.2	C.1	C.ar	C.cat	N.3	N.2	N. 1	N.ar	N.am	N.pl3	N.4	O.3	0.2	0.CO2	S.3	S.2	S.O	S.O2	P.3	H	F	CL	BR	I	NO2	NO	CN	CO	SO2	SO	OH
Level 0							1																								
Level 1		2																													
Level 2		1									1										2										
Level 3	1																				1										
Level 4	1																				2										
Level 5						1															1										

FIGURE 6





	No. of atoms of certain type at each level			
	N.ar	C.ar	H	N.pl3
Level 0	1			
Level 1		2		
Level 2		3	1	
Level 3		3	2	
Level 4		1	3	
Level 5				1

Construction of the hierarchical tree from one example, 6-amino quinoline

FIGURE 7

pK_a of Bases (384)

PC = 5, r^2=0.922 q^2=0.832, Std. Err = 0.89

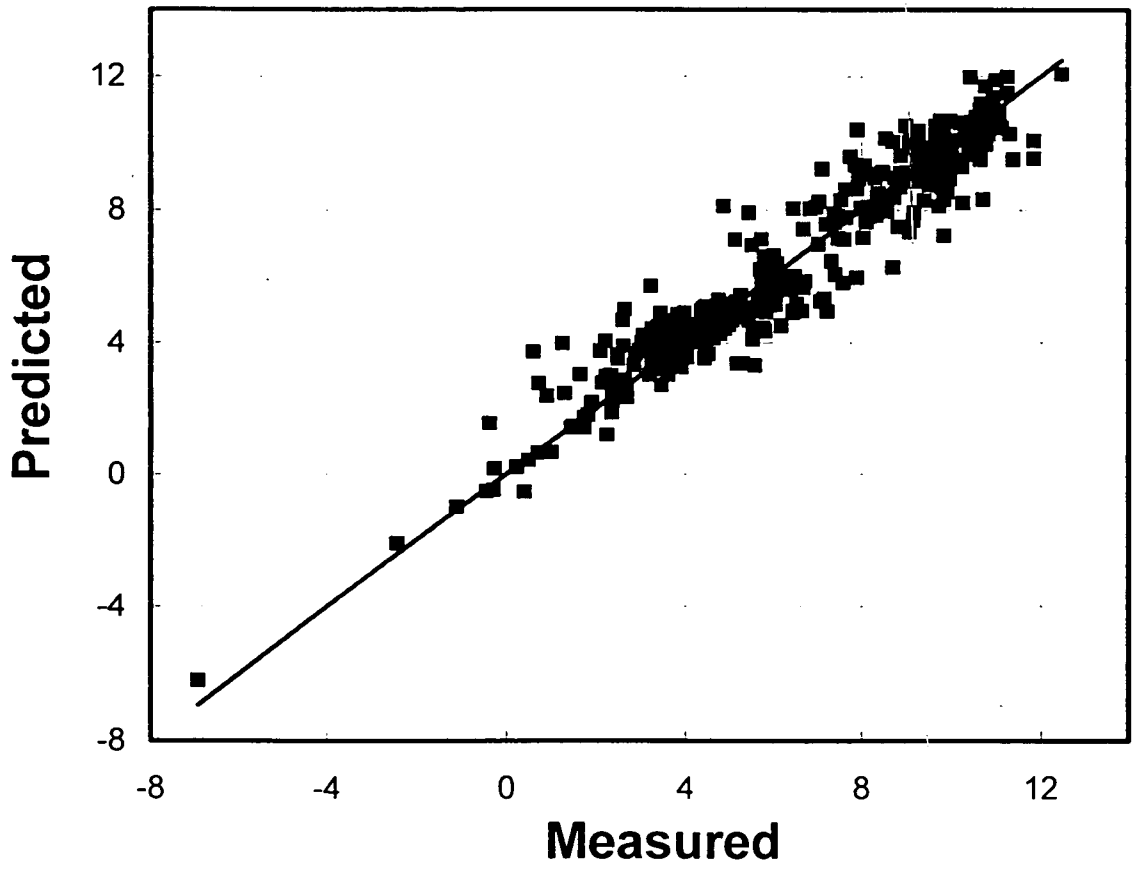


FIGURE 8

pK_a of Acids (646)

PC = 6, r^2 = 0.927, q^2=0.792, Std. Err=0.77

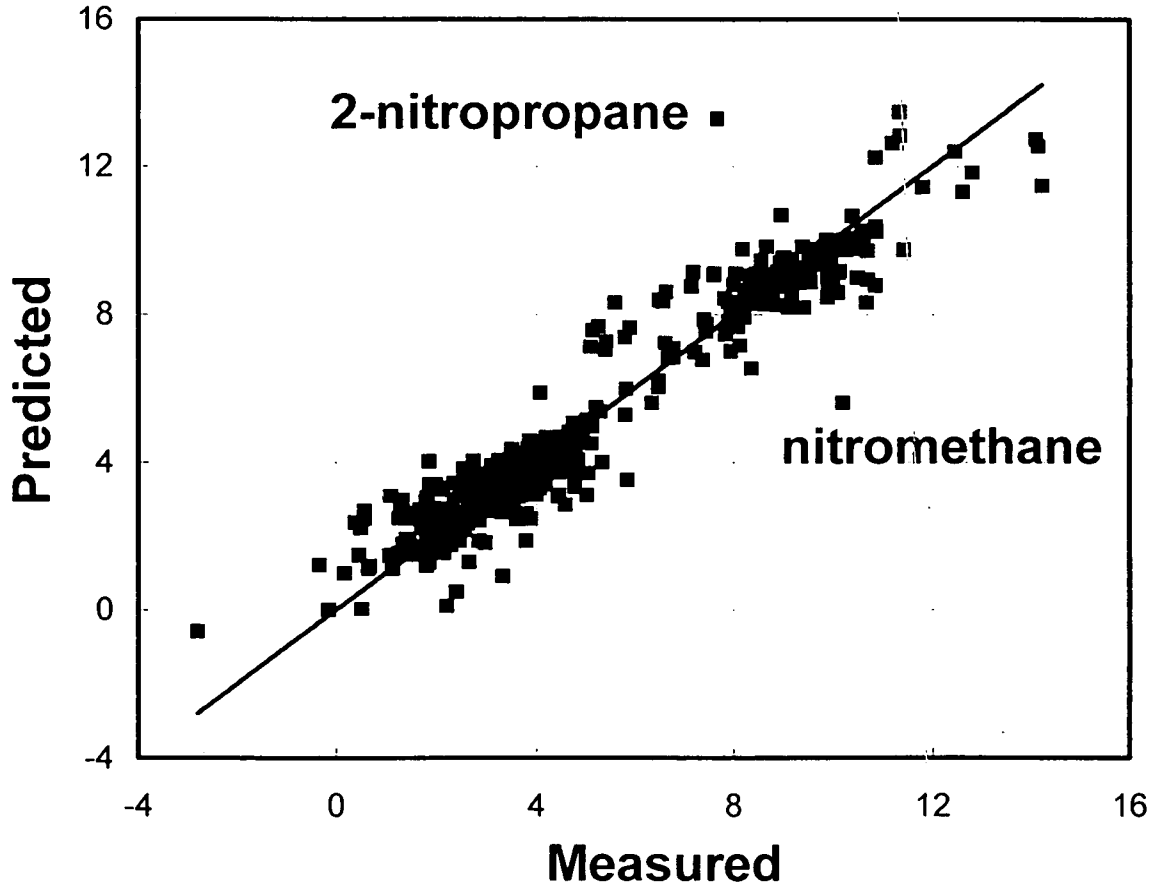
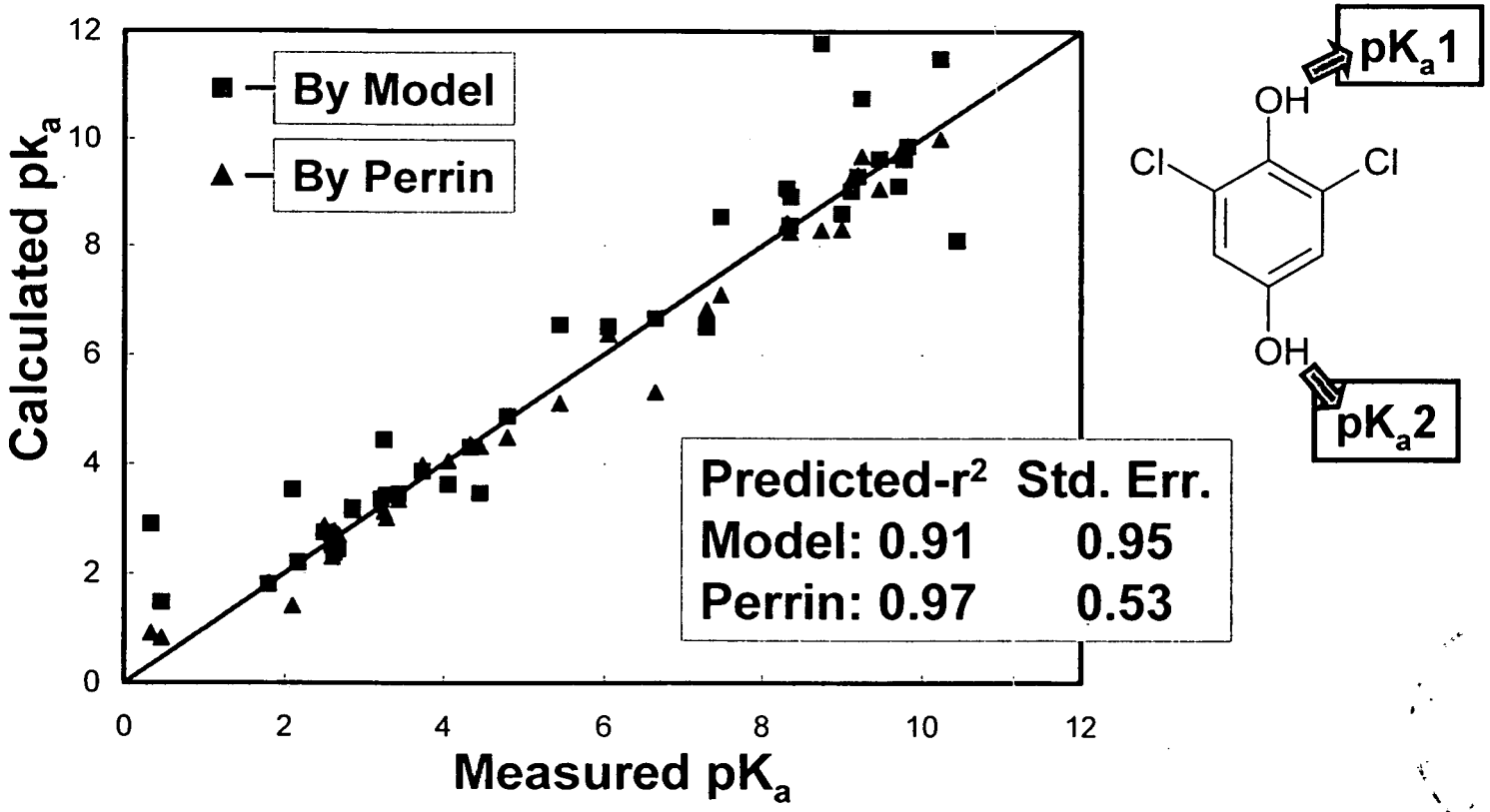


FIGURE 9

Prediction on Molecules

- 37 acid and base molecules selected by Perrin* *et al.* as examples, and were not included in the training set
- For molecules containing multiple ionizable centers the model correctly predicted the order for all of them, ie



*Perrin DD, Dempsey B, Serjeant EP, *pKa prediction for organic acids and bases*, Chapman and Hall, London, 1981.